ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

Sentiment Analysis and Ranking of the Products based on the Reviews

Monika Rani HG[1], Sourabh Jain[2], Mohd Danish Ahmad[3], Mohit Sehrawat[4], Rishabh Raj[5]

Assistant Professor, Dept. of Computer Science & Engineering, Sir MVIT, Bangalore, India. [1]

B. E Scholars, Dept. of Computer Science & Engineering, Sir MVIT, Bangalore, India. [2 to 5]

ABSTRACT:The world wide web can be viewed as a repository of opinions from users spread across various websites and networks, and today's netizens look up reviews and opinions to judge commodities, visit forums to debate about events and policies. With this explosion in the volume of and reliance on user reviews and opinions, manufacturers and retailers face the challenge of automating the analysis of such massive amounts of data (user reviews, opinions, sentiments). Armed with these results, sellers can enhance their product and tailor experience for the customer. Similarly, policy makers can analyse these posts to get instant and comprehensive feedback. Or use it for innovative ideas that democratize the policy making process. This paper is the outcome of our research in gathering opinion and review data from popular portals, e-commerce websites, forums or social networks; and processing the data using the rules of natural language and grammar to find out what exactly was being talked about in the user's review and the sentiments that people are expressing. Our approach diligently scans every line of data, and generates a cogent summary of every review (categorized by aspects) along with various graphical visualizations. A novel application of this approach is helping product manufacturers or the government in gauging response.

KEYWORDS: Sentiments; analysis; software classification; e-commerce websites; reviews; ranking.

I. INTRODUCTION

Today the sheer volume of data being that is generated ever is enormous and making any sense out of that data is a tedious task. However, constant efforts and research in this area have led the automation of the process to some extent. With this project, we aim to further this automation process. Using a combination of data aggregation techniques, NLP, linguistic analysis and popular visualization techniques we generate visually appealing and easy to understand graphs which provide summarized feedback. This is done by performing detailed sentiment analysis on the data. The fields of opinion mining and sentiment analysis are distinct but deeply related. Opinion mining focuses on polarity detection [positive, negative or neutral] whereas sentiment analysis involves emotion recognition. Because detecting the polarity of text is often a step, in sentiment analysis.

In the proposed approach, the algorithm will first get the reviews of products from the given URL and then parse the reviews clean them. Find the positive and negative polarity for each review against the product. The product is again rated on the various attributes namely Screen, Phone, Price, Speaker, Battery, Camera & Quality and then provide the overall sentiment distribution of product.

II. RELATED WORK

Summarizing sentiment and summarizing sentiment by extracting and aggregating sentiment over rateable aspects has been a very active area of research recently. In the previous approach the user demonstrated a multi-step methodology that combines cutting edge techniques in multi-document summarization, polarity classification and sentiment analysis. IT automatically synthesizes, filters, and summarizes user product reviews into a compact list of positive and negative opinions as expressed in the reviews.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

Disadvantages of Previous Approach:

[1] The previous approach just finds reviews of previous product manually by collecting data in the form of an excel file.

[2] The previous approach will produce and classify the reviews as positive and negative but does not consider any attributes to get the best product.

III. PROPOSED METHOD

In the current approach, the application provides the flexibility to collect offline reviews as well as online reviews from Flipkart, Amazon, Snapdeal, Walmart, etc. using Web Crawler algorithm. After collecting the reviews, a series of data mining algorithms like data cleaning, tokenization and frequency computation are performed.

The current approach also provides the concept of features. The features can be described as follows:

1) Camera
2) Memory
3) Battery
4) Touch
5) Sound
6) Screen

The feature based frequency is also computed for all the collected reviews.

The next sequence is to determine the sentiments namely Positive Polarity, Negative Polarity and Neutral Polarity by dividing the reviews into sentences and then computing the 3 polarities per review and per feature, also the polarities across reviews are added together to obtain Product based polarity. Once the product based polarity and feature based frequency are obtained then the products are ranked based on Positive Polarity Maximum, Negative Polarity Minimum and Neutral Polarity Maximum.

IV. ALGORITHM

The Methodology of the algorithm can be described as follows:

1. Collection of reviews for the given Product and URL.
2. Break the reviews into a sequence of statements.
3. For each review and the corresponding statement, we need to perform the POS tagging to get POS tagged statements.
4. For each statement, the verbs are found out and checked whether the verbs name is in the set {Screen, Phone, Price, Speaker, Battery, Camera, Quality}.
5. For each statement and verb, the adjectives are found and then they are compared against a set of features across products.
6. Feature Extraction Matrix is generated and product recommendations are generated based on the user searched feature. The FEM (feature extraction matrix) will help further in counting the number of positive and negative reviews for each feature.

V. PHASES(MODULES) OF SIMULATION

Review Collection

Offline Phase

Input- Review Description, Product Name.

Output – Review Stored at the Data source.

Online Phase

Input – URL of review, Product Name, Site-Amazon or Flipkart.

Output – Collected reviews from Amazon or Flipkart and stored in the format of Reviews in the data store.

Data Cleaning

Input – Collected Reviews and Stopwords.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

Output – Cleaned reviews which does not contain any un meaningful words.

Frequency Computation

Input – Cleaned Reviews.

Output – Tokens Formation and Frequency Computation per review.

Feature based Frequency Computation

Input- List of Tokens, Review IDs, Product IDs.

Output – Computation of Frequency across all products and all reviews per feature.

Sentiment based Computation

Input – Reviews.

Output- Positive, Negative and Neutral Sentiments.

Feature Extraction Matrix (FEM) generation algorithm

Input – Sentiment.

Output – FEM matrix which has one row per product and one column per feature.

Ranking Algorithm

Input – FEM matrix and Search Query.

Output – List of Products based on highest value of FEM for searched features.

VI. SIMULATION RESULTS

Simulation results shows the pictorial representation of different phases of the simulation. Offline reviews are submitted by providing the review details and product details as shown in the fig1.

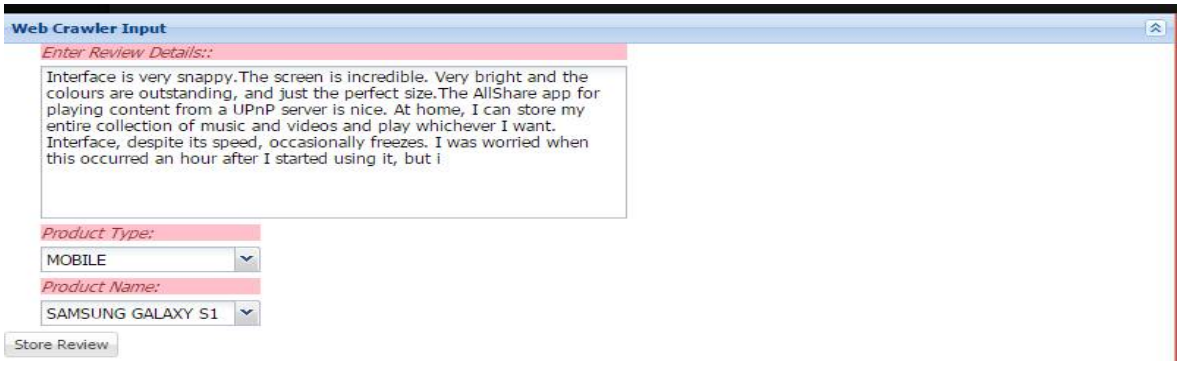

Fig. 1. Review Submission

Online review collection also can be done by providing the URL of the site, using web crawler algorithm to fetch the reviews about the products from the website. The same has been simulated and got the results as shown in the fig2.

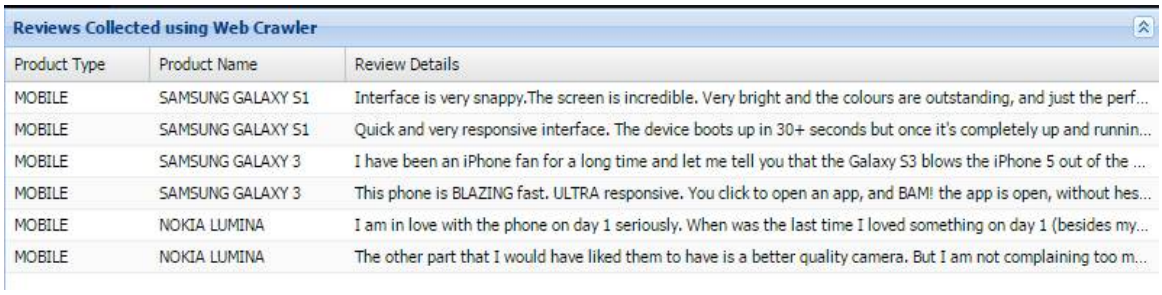

Fig. 2. Online Review Collection using Web Crawler Algorithm

Extracting the important keywords from the reviews and removing unimportant words done using data cleaning process.The cleaned reviews are as shown in the fig 3.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

Reviews Output

Clean ID	Review ID	Product ID	Product Type	Review Details
1	1	1	1	interface snappy screen incredible bright colours outstanding perfect size allshare app playing content u...
2	2	1	1	quick responsive interface device boots seconds s completely running seconds starting services media sc...
3	3	2	1	iphone fan long time galaxy s blows iphone water skeptical galaxy s compare iphone s competition galax...
4	4	2	1	phone blazing fast ultra responsive click open app bam app open hesitation love screen read read review...
5	5	3	1	love phone day time loved day girl thing improved transitioning iphone app market lot financial tools llik...
6	6	3	1	part quality camera complaining end love lumia app market phone

Fig. 3. Cleaned Reviews

The frequency computation and tokenization is performed on cleaned reviews.Fig4 shows the tokenized frequency matrix.

View Frequency Output

Review ID	Product ID	Product Type	Token Name	Frequency ID	Frequency
1	1	1	interface	1	2
1	1	1	snappy	2	1
1	1	1	screen	3	1
1	1	1	incredible	4	1
1	1	1	bright	5	1
1	1	1	colours	6	1
1	1	1	outstanding	7	1
1	1	1	perfect	8	1
1	1	1	size	9	1
1	1	1	allshare	10	1
1	1	1	app	11	1
1	1	1	playing	12	1

Fig. 4. Tokenised Frequency Matrix

Fig 5 shows the feature extraction results and frequency computation per feature per product.

Feature Output

Feature ID	Review ID	Product ID	Product Type	Feature Type	Feature Based Freq
1	1	1	1	CAMERA	0
2	1	1	1	MEMORY	0
3	1	1	1	BATTERY	0
4	1	1	1	TOUCH	0
5	1	1	1	SOUND	0
6	1	1	1	SCREEN	1
7	2	1	1	CAMERA	0
8	2	1	1	MEMORY	0
9	2	1	1	BATTERY	0
10	2	1	1	TOUCH	0
11	2	1	1	SOUND	0
12	2	1	1	SCREEN	0

Fig. 5. Feature Vector Matrix

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

Fig6 shows categorizing the reviews into positive, negative, neutral.

Total Polarity Output

Product ID	Product Type	Positive Rating	Negative Rating	Neutral Rating	Feature Type	Total Feature
1	1	0	0	2	CAMERA	0
1	1	0	0	4	MEMORY	0
1	1	0	0	6	BATTERY	0
1	1	0	0	8	TOUCH	0
1	1	0	0	10	SOUND	0
1	1	1	0	11	SCREEN	1
2	1	0	0	2	CAMERA	0
2	1	0	0	4	MEMORY	0
2	1	0	0	6	BATTERY	0
2	1	0	0	8	TOUCH	0
2	1	0	0	10	SOUND	0
2	1	0	0	12	SCREEN	1

Fig. 6. Product based polarity

Fig7 shows the graphical representation (product id v/s neutral rating) of feature (e.g.: Battery) neutral polarity.

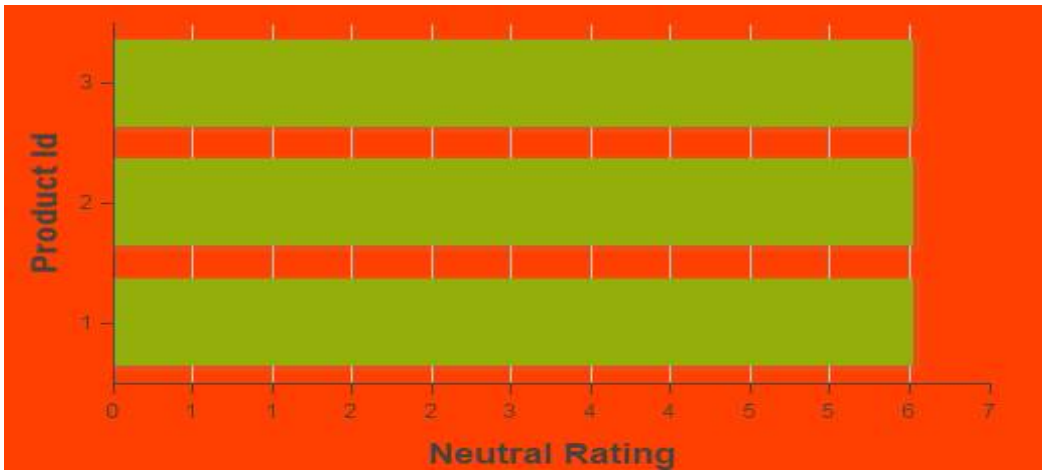

Fig. 7. Neutral Polarity Graph for Feature – Battery

Fig8 gives the UI for the user to search product (e.g.: Mobile) with feature.

Fig. 8. Mobile Search

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

Fig9 Displays the list of products according to their ranks(FEM) base on the user search query.

Product Name	Product Id	Product Type	Total Feature	Total Negative	Total Neutral	Total Positive
NOKIA LUMINA	3	0	0	0	18	0
SAMSUNG GALAXY 3	2	0	0	0	12	0
SAMSUNG GALAXY S1	1	0	0	0	6	0

Fig. 9. Product Ranking based on the Searched Content

VII. CONCLUSION AND FUTURE WORK

As size of information present on the internet has taken a shape of the giant, it has become a necessity to increase the efficiency of the search engines. Web mining is aiming in this direction. In this work, we have done both feature-based and based on negative, neutral and positive polarity. In future----

1) More Number of products can be taken into consideration, so that more products can be recommended.
2) More Number of features can be taken into consideration to suggest more unique and good products.

REFERENCES

1. F. Zhu and X. Zhang, "Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics," Journal of marketing, vol. 74, no. 2, pp. 133–148, 2010.
2. Q. Ye, R. Law, and B. Gu, "The impact of online user reviews on hotel room sales," International Journal of Hospitality Management, vol. 28, no. 1, pp. 180–182, 2009.
3. Q. Roller, "Quote roller reviews, pricing comparison of alternatives," GetApp, vol. 02-, 2015. [Online]. Available: http://www.getapp.com/ sales- software/a/quote- roller/
4. N. Mishra and C. Jha, "Classification of opinion mining techniques," International Journal of Computer Applications, vol. 56, no. 13, pp.1–6, 2012.
5. S. Model, "Software as a service (saas) from the gartner IT glossary," Gartner IT Glossary, vol. 04-, 2012. [Online]. Available: http://www.gartner.com/it-glossary/software- as-a- service-saas/
6. B. Pang and L. Lee, "Opinion mining and sentiment analysis," Foundations and trends in information retrieval, vol. 2, no. 1-2, pp. 1–135, 2008.
7. D. Song, H. Lin, and Z. Yang, "Opinion mining in e-learning system," in Network and Parallel Computing Workshops, 2007. NPC Workshops. IFIP International Conference on. IEEE, 2007, pp. 788–792.
8. E. Marrese-Taylor, J. D. Vela´squez, F. Bravo-Marquez, and Y. Matsuo, "Identifying customer preferences about tourism products using an aspect-based opinion mining approach," Procedia Computer Science, vol. 22, pp. 182–191, 2013.
9. S. F. Sayeedunnissa, A. R. Hussain, and M. A. Hameed, "Supervised opinion mining of social network data using a bag-of-words approach on the cloud," in Proceedings of Seventh International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012). Springer, 2013, pp. 299–309.
10. N. Bharathi, G. Brindha, B. Santhi, and P. Neelamegam, "Integration of fpga, cloud and opinion mining," Journal of Applied Sciences, vol. 14, no. 14, p. 1653, 2014.
11. M. Cooke and N. Buckley, "Web 2.0, social networks and the future of market research," International Journal of Market Research, vol. 50, no. 2, p. 267, 2008.
12. Y. Chen and J. Xie, "Online consumer review: Word-of-mouth as a new element of marketing communication mix," Management Science, vol. 54, no. 3, pp. 477–491, 2008. [Online]. Available: http://pubsonline.informs.org/doi/abs/10.1287/mnsc.1070.0810
13. B. Liu, "Sentiment analysis and opinion mining," Synthesis Lectures on Human Language Technologies, vol. 5, no. 1, pp. 1–167, 2012.
14. L. Zhuang, F. Jing, and X.-Y. Zhu, "Movie review mining and summarization," in Proceedings of the 15th ACM international conference on Information and knowledge management. ACM, 2006, pp. 43–50.
15. A. Minanovic, H. Gabelica, and Z. Krstic, "Big data and sentiment analysis using knime: Online reviews vs. social media," in Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014 37th International Convention on. IEEE, 2014, pp. 1464–1468.
16. V. Pekar and S. Ou, "Discovery of subjective evaluations of product features in hotel reviews," Journal of Vacation Marketing, vol. 14, no. 2, pp. 145–155, 2008.
17. W. Kasper and M. Vela, "Sentiment analysis for hotel reviews," in Computational Linguistics-Applications Conference, vol. 231527, 2011, pp. 45–52.
18. Cloudreviews, "com, 'Cloud Reviews — Cloud Hosting — Managed Cloud — Cloud storage Apps'," [Online]. Available: http://www.cloudreviews.com/. [Accessed: vol. 02-, 2015. [Online]. Available: http://www.cloudreviews.com/
19. P. Simone Guercini, F. Misopoulos, M. Mitic, A. Kapoulas, and C. Kara- piperis, "Uncovering customer service experiences with twitter: the case of airline industry," Management Decision, vol. 52, no. 4, pp. 705–723, 2014.
20. B. Liu and L.Zhang, "A survey of opinion mining and sentiment analysis," in Mining text data. Springer, 2012, pp. 415–463.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

21. E. H. Thomas and N.Galambos, "What satisfies students? mining student-opinion data with regression and decision tree analysis," Re- search in Higher Education, vol. 45, no. 3, pp. 251–269, 2004.
22. D. Lee, O.-R. Jeong, and S.-g. Lee, "Opinion mining of customer feedback data on the web," in Proceedings of the 2nd international conference on Ubiquitous information management and communication. ACM, 2008, pp. 230–235.
23. Lexalytics, "com, 'SaaS and on-premise text analytics — lexalytics'," [Online]. Available: https://www.lexalytics.com/. [Accessed: vol. 02-, 2015].
24. Q. Ye, Z. Zhang, and R. Law, "Sentiment classification of online reviews to travel destinations by supervised machine learning approaches," Expert Systems with Applications, vol. 36, no. 3, pp. 6527–6535, 2009.
25. R. Platform, "Rapidminer - 1 open source predictive analytics platform," RapidMiner, vol. 02-, 2015.